processes, shown in **Fig. 4**, are on the whole 'unnatural' at best. While the configuration of points does not strongly suggest any particular set of points to be a cluster, it does suggest symmetry. The lack of symmetry in the results in Fig. 4 tends to indicate the sensitivity of the algorithms to small local perturbations. The distance matrix of ranks was clustered by our proposed method using $k = 3$. **Fig. 5** shows the computer output of the cluster candidates and the corresponding node values and isolation indices. Examination of the isolation indices reveals that clusters 8 and 9 may reasonably be considered as 'real'. Under the probability model studied in Ling (1971b), which is not quite appropriate for the data under consideration but nevertheless offers a rough guide to the assessment of the significance of the clustering indices, these clusters are seen to be statistically significant. **Fig. 6** shows the tree diagram for the results of clustering. The symmetry of the configuration, which was not brought out by other methods examined, is easily seen here.

5. Conclusion

The clustering method proposed in this paper represents a modest attempt by this author to introduce rigour into the subject of cluster analysis to make it logically more satisfying than it is in its current state. For problems of limited size as those treated in this paper ($n \leqslant 100$), the approach and results seem encouraging. The study of probability theory (exact and approximate) associated with the clustering indices is possible, though difficult. Even without the support of probability theory, one can state precisely what mathematical properties each cluster possesses—a desirable trait shared by very few existing methods. The major limitation of our method is its impracticality for large size problems, say $n > 500$, when both the computer storage and the computing time requirements quickly become prohibitive. On the whole, this author believes the approach of explicitly defining what constitutes clusters worthy of further exploration and hopes that others will devote more of their attention to the fundamental question: 'What is a cluster?'

References

BONNER, R. E. (1964). On some clustering techniques, *IBM Journal of Research and Development*, Vol. 8, pp. 22-23.

GOWER, J. C., and ROSS, G. J. S. (1969). Minimum spanning trees and single linkage cluster analysis, *Applied Statistics*, Vol. 18, pp. 54-64.

HARTIGAN, J. A. (1967). Representation of similarity matrices by trees, *Journal of the American Statistical Association*, Vol. 62, pp. 1140-1158.

JARDINE, N., and SIBSON, R. (1968). The construction of hierarchic and non-hierarchic classifications, *The Computer Journal*, Vol. 11, pp. 1177-1184.

JOHNSON, S. C. (1967). Hierarchical clustering schemes, *Psychometrika*, Vol. 32, pp. 241-254.

LANCE, G. N., and WILLIAMS, W. T. (1966). Computer programs for hierarchical polythetic classification (Similarity analyses), *The Computer Journal*, Vol. 9, pp. 60-64.

LANCE, G. N., and WILLIAMS, W. T. (1967a). A general theory of classificatory sorting strategies I. Hierarchical systems, *The Computer Journal*, Vol. 9, pp. 373-380.

LANCE, G. N., and WILLIAMS, W. T. (1967b). A general theory of classificatory sorting strategies II. Clustering systems, *The Computer Journal*, Vol. 10, pp. 271-277.

LING, R. F. (1971a). Cluster analysis, unpublished Ph.D. dissertation, Department of Statistics, Yale University.

LING, R. F. (1971b). Probability theory of cluster analysis, presented at the Second Annual Meeting of the Classification Society (North American Branch), London, Canada, to be published in *Journal of the American Statistical Association*.

MCQUITTY, L. L. (1960). Hierarchical linkage analysis for the isolation of types, *Education and Psychological Measurement*, Vol. 20, pp. 55-67.

SHEPHERD, M. J., and WILLMOTT, A. J. (1968). Cluster analysis on the Atlas computer, *The Computer Journal*, Vol. 11, pp. 57-62.

SIBSON, R. (1971). Some observations on a paper by Lance and Williams, *The Computer Journal*, Vol. 14, pp. 156-157.

SNEATH, P. H. A. (1957). The application of computers to taxonomy, *Journal of General Microbiology*, Vol. 17, pp. 201-226.

SOKAL, R. R., and SNEATH, P. H. A. (1963). *Principles of Numerical Taxonomy*, San Francisco: W. H. Freeman and Co.

WARD, J. H. (1963). Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, Vol. 58, pp. 236-244.

WISHART, D. (1969). An algorithm for hierarchical classification, *Biometrics*, Vol. 25, pp. 165-170.

Correspondence

To the Editor
The Computer Journal

Sir

Mr. A. J. Finn suggests (this *Journal*, Vol. 15, No. 1, p. 12) an extension to FORTRAN which would permit generalised iteration parameters in a DO statement.

Unfortunately, his example includes an arithmetic expression; if he really intends to permit them, his proposed syntax is ambiguous. The statement

DO n I = /n1, n2/n3/n4, n5, n6/I + (integer variable or expression), (Boolean variable or expression)/

was intended to imply the following values for the induction variable I:

n1 to n2 by 1
n3
n4 to n5 by n6
I + (integer expression)
all WHILE (Boolean expression).

Several other meanings are possible, such as

1. n1 to n2/n3 by 1
   n4 to n5 by n6 etc.
2. n1 to n2 by 1
   n3/n4 to n5 by n6 etc.
3. n1 to n2/n3 by 1
   n4 to (n6/I) + expression by n5 etc.

While I favour the notion of a Boolean expression as the fourth DO parameter to define a WHILE condition, it would appear that the present syntax of the DO statement is singularly ill-adapted to the addition of iteration lists.

Yours faithfully,

J. R. EHRMAN

Stanford Linear Accelerator Centre
Stanford University
P.O. Box 4349
Stanford
California 94305
USA

12 May 1972